AN ENHANCED DATA TRANSFORMATION FRAMEWORK FOR THE SONIFICATION OF SIMULATED RIGID BODIES

Rhys Perkins

Anglia Ruskin University,
Department of Music and Performing Arts,
Cambridge, CB1 1PT, UK
rhysperkins@rhysperkins.com

ABSTRACT

While simulated rigid bodies hold a wealth of information that can be understood through sound (Table 1), current interpretive methods will often overlook important features of their data. This proves to be detrimental when placing the same data in the context of an auditory display where the user might wish to analyse or express specific dimensions under a range of circumstances. The following investigation describes a framework for a model-induced parameter mapping technique which allows for an explicit level of control over the flow of information, supported by a number of key conditions in both the auditory and visual channels. Given that the formative decisions behind this design tailors the data to meet the purposes of sonification, the user is presented with a viable alternative that overcomes a number of limitations inherent to employing a more conventional physical modelling approach.

Property	Type	Example
Position	Vector	(11.2, 4.3, 1.0)
Orientation	Quaternion	(1.0, 0.0, 0.0, 0.0)
Angular Velocity	Vector	(0.0, -0.8, 0.0)
Linear Velocity	Vector	(0.2, -9.4, 0.4)
Angular Momentum	Vector	(0.8, 34.0, 4.8)
Linear Momentum	Vector	(9.0, 1.5, 12.5)
Mass	Scalar	1.8
Scale	Vector	(1.0, 0.5, 1.0)

Table 1: Rigid body data - Common properties of a simulated rigid body alongside their corresponding type

1. INTRODUCTION

The simulation of rigid body dynamics has become more widespread in recent years. This is partly due to the increasingly mandatory integration of a physics engine within cross-platform game engines, such as Unity [1], to facilitate the rendering of simulated objects with real-time interaction and audible response on mobile and desktop devices. In spite of this, it is apparent that the formative decisions behind the ongoing development of this gaming architecture are not focused on auditory display and are instead concerned with realism and immersion. This is particularly evident when reviewing the efforts to invest simulated rigid bodies with the attributes of sound. In that regard, the predominant approach is to accept increasingly accurate physical modelling methods as a means to produce convincing audio which accompanies the simulation of physical events [2][3]. However, it is suggested here that this methodology is not necessarily flexible enough for the purposes of sonification due to its strict simulation of physical ties between mechanical and acoustic systems. Furthermore, while there are several examples which demonstrate that the sonification of interactive rigid bodies within the context of game engine could provide a feasible alternative [4][5], they do not provide the theoretical evidence to support such a case.

The following investigation concentrates on the data transformation process in more detail by presenting the theoretical framework which underpinned the formative decisions behind the corresponding component of Mhyusics Auditory Display [6]. It begins by considering how rigid body data can be interpreted effectively through the visual channel before introducing a framework which can simultaneously interpret the same data through sound while continuing to be supported by its visual counterpart. This identifies a synergistic approach to the presentation of information which is then compared to physical modelling in order to highlight potential benefits for the user. Audiovisual examples which demonstrate this technique can be found online [7].

2. RIGID BODY DATA AND THE VISUAL CHANNEL

As one embarks on the design process of a sonification, understanding the chosen data set becomes a significant area of interest. This can bring with it some considerable challenges as Worrall [8] acknowledges when he describes it as being the first major bottleneck:

> "In data sonification, whilst the input data can be thought of as eventually controlling the sound rendering, the transformations it has to undergo in the interim can be considerable. Such data processing can reasonably include multidimensional scaling, filtering and statistical analysis which itself may itself become the subject of sonification."

According to Hunt and Hermann [9] we face at least two fundamental problems in the domain of data exploration. First, "the data often inhabit a high-dimensional data space that is very different from the 3D space we are familiar with" and secondly, "these data spaces have no intrinsic means of interacting with them". Both of these issues can be addressed when the data is derived from a simplified analogy of physical objects and placed in the framework of a game engine.

In most game engines the correlation between data and video has been well established by allocating polygonal models to depict rigid bodies in a faux, three-dimensional Euclidean space (Figure 1). This proves to be pertinent when considering that Hunt and Hermann have proposed that "model-based approaches may offer the chance to bind together different modalities into a useful whole, both for display and interaction purposes". Their method requires the designer to map the data to an intermediary model rather than directly to sound. Given the similarity in approach to interpreting data, we can regard the polygonal models in this context as one of the most fundamental forms of model-based sonification (MBS) [22]. This brings with it the many of the benefits of MBS but instead places the intermediary model in the virtual environment, allowing it to be viewed from any number of angles and at an arbitrary distance. In that respect, the models not only contextualise the data and reduce its complexity by means of pre-interpretation, they also instantiate an elementary means for direct manipulation which permits both the user and an audience to perceive the direct relationship between action and effect.

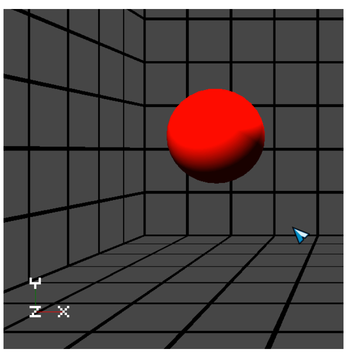

Figure 1: Example of a simulated rigid body displayed as a red polygonal model in an interactive three-dimensional environment

2.1. Characterising the data through tasks

In order to understand the underlying data for the purposes of an AD, Barrass [10] proposes the combining of task analysis and data characterisation methods. His design approach draws attention to how an AD will be used through the identification of task related goals. In this sense, Hunt and Hermann [9] demonstrated that we perform tasks and data analysis through our everyday interaction with objects. It would therefore be reasonable to assume that we can apply the same techniques to their simulated equivalent when they are portrayed via a common channel of perception. This was substantiated by Blackwell [11] who gave the simulation approach to metaphor [12] as an example to suggest that "safety for the designer lies in mimicking a nondigital artifact in order that the user's actions are predictable".

When placing an AD in the context of a game engine, we can allow computer-mediated tasks to be determined by the manipulation of the visually simulated bodies. In turn, these actions are directed by gestures that are mediated through our hand movements. Continued investigation by Cadoz [13] has observed several gestural methods for the performing of functions in a virtual space. He divided the approach for accomplishing tasks involving simulated bodies into two broad typologies. The first, described as the ergotic function, has similar connotations to that of direct manipulation entailing "exchanges of energy between a human body and material objects or a material environment". Conversely, the second, or non-ergotic, function occurs when "forces, displacements or exchanges of energy are involved only and exclusively with the body of the subject". It is this second type that leads us to consider the influence of the computer on gestural interaction.

Kojs [14] digitised human gestures which could be performed on simulated entities incorporating the rigours of physical modelling synthesis when formulating what he called action-based music. While his cyberactions all derived from the same mechanical principles, they were extended by the computer to create increasingly incongruent actions resulting in unpredictable energy flow and extrinsic sonic response. The performance of these actions took place in a virtual environment [15] which allowed for the reconfiguring of physically modelled entities, such as exciters and resonators, to form new actions and galvanise the creation of novel compositions in cyberspace. These achievements serve to indicate that object related tasks within similar mental models are still open to user interpretation and encourage the prospect of further work. Collateral evidence of this prospect emerges from Blackwell [11] who concludes that when depending on metaphor as a visual communication channel "the design ideal in this case is to provide effective access to this information, while allowing expert users to bypass the metaphor if they already have the necessary information". Certainly, the display of physically simulated models brings with it a sense of familiarity that invites us to rely on our common sense knowledge when interpreting and communicating ideas. However, the typical physics engine also allows us to circumvent the standard model behaviour.

Let us consider that in reality we witness objects governed by the laws of our environment, the actions of which are exemplified by the physics engine as the dynamic rigid body. Each time the engine updates, it recalculates the body's world transform. This transform is specified through its scale, rotation and translation, which changes in response to the forces exerted from environmental factors to give the sense of realism that we come to expect from its visual behaviour. Kinematic bodies also comprise the same world transform but are capable of ignoring environmental factors, such as gravity and resistive forces against motion, thus denying our expectations. They permit a one-way transfer of energy when interacting with dynamic bodies by admitting an influence upon them but remaining unresponsive in return. Both of these core types can be animated by the user and the computer, but only the dynamic objects have the capacity to act after a gesture has concluded. This poses a number of conceivable scenarios where kinematic objects can be employed to restrict or guide gesture and motion relating to their dynamic counterparts.

Additionally, changing between the two types during the course of a simulation can incite abrupt changes in movement and drastically alter the energy flow of neighbouring objects.

2.2. Summary

By relying on established visual relationships pertaining to the real world we are looking to create an ideal foundation for sonifying rigid bodies. This is helped, in part, by working solely with a single data set whose properties are constrained by the simulation algorithms of rigid body dynamics. The entities of the natural world, which incorporate the behaviour that these algorithms seek to replicate, have a longstanding synergy with humans. This would suggest that less mental bandwidth is required to visually comprehend tasks and events. Instead, the attentive capacity of the user can focus on the audio, along with its governing mapping configurations, as a means of encouraging sonic exploration. Worrall [8] characterises this situation as "data SONIFICATION" where "the primary focus is on sound rendering whilst input data is constrained so it can be dealt with adequately by the rendering software". This avoids the alternative method, known as "DATA sonification", which has an "emphasis on data-processing tools at the expense of sound rendering flexibility".

3. RIGID BODY DATA AND THE AUDITORY CHANNEL

At this point we have established that the connection between data and video relies on a polygonal model to encapsulate the data set of a simulated rigid body which, in turn, facilitates inherent methods for task analysis. However, it can be argued that the process of task analysis may be enhanced by the introduction of sound, which can coincide with the visual modality to reveal the current state of each model.

Hermann [16] illustrated that the majority of our everyday tasks involving physical objects are accompanied with sonic feedback, which acts to aid us in their processing. This led to the conception of analytical everyday listening [17] which suggested that we could analyse the properties of an object under investigation due to the unique sound each interactive task procured. Both these concepts were based on the understanding that "the meaning of an acoustic event is primarily rooted in conveying information about important physical properties of an object or process". This proves to be of some importance when considering their statement that:

> "compared to other contexts, the context given by physical laws was stable all the time, so that evolution had ample time to adapt our brains extremely well to the ways how physics links sounds and their causes. This is reflected in a number of rather "universal" relationships that are deeply engrained in the way we - usually subconsciously - pick up meaning from sound events."

Indeed, the idea of invariance between sound and data also supports the act of everyday listening [18] which upholds the proficiency of the auditory modality for the comprehension of object data arising from physical events, rather than tasks. Nevertheless, while both these modes of listening are relevant to rigid body interactions, they tend to rely on known associations that have been ingrained through experience. Consequently, they may be less suited to more variable cases where the connections between information and sound are yet to be forged.

3.1. Data familiarisation through gestalt principles

Kramer [19] understood that the user will not necessarily deem data and audio to be inherently linked during their initial experience of a typical AD. He argues that when placing the user in this situation they will automatically begin to identify structures and patterns which emerge from a more complex array of sounds. This perceptual event is regarded by Kramer as an auditory gestalt, which implies a prominent feature of the underlying data and is therefore a crucial factor in the forging of a link between data and sound:

> "As we learn to use a representation technique and become familiar with its gestalts, we recognize gestalts as signatures of specific events. If the universe of possible events is sufficiently limited, we in effect learn the "language" of the display, wherein each class of gestalts symbolizes a general category of data event or system state."

His argument is supported by the theory of auditory scene analysis [20] which distinguishes a number of grouping cues to fuse and segregate streams of sound as a means of identifying their source. Likewise, these cues are instinctively employed in our everyday listening and often resemble those classified by the gestalt principles for perceptual organisation. For instance, both the law of similarity and fusion cues can refer to common audio elements such as timbre and frequency, while the law of common fate and segregation cues can be influenced by periodic components such as those with amplitude modulation. Both sets of principles can be applied to any AD that relies on the user's discretion to form their own connections as their brain naturally learns to recognise circumstances which associate certain sound states with equivalent structural states. Over time, the user will find that less cognitive effort is required to link the data to the sonic representation and that the audio becomes more efficient in categorising underlying data events. Kramer [19] compares this process to musical training where "a student learns to identify certain chords or intervals from what was previously a sea of notes rising and falling".

Hermann and Ritter [17] regarded auditory gestalt perception as an important mode of listening which complemented their idea of analytical everyday listening. They asserted the concept of auditory gestalts as analogous to visual gestalts calling them "a subset of acoustical elements perceptually bound together into a 'unit' as a result of a particular coherence, characterized by one of the 'gestalt laws'". With this in mind they put forward that sonification models can support both their acquisition and the learning process "by supplying an invariant process to be used in the same manner for very different data sets". The same can be said for the polygonal models in this framework, which reflect the internal parameters of a rigid body in a consistent manner. Accordingly, the search for gestalts becomes more efficient as the visual evidence not only serves to reinforce where gestalts occur, but distinct model behaviour and interaction can predicate their emergence.

3.2. A foundation for data transformation

De Campo [21] understood that "the most general task in data sonification designs for exploratory purposes is to detect auditory gestalts in the acoustic representation, which one assumes correspond to any patterns and structures in the data one wants to find". In order to ascertain guidelines for the conception of these auditory gestalts he produced what he termed as the Sonification Design Space Map (SDSM). The groundwork for its design involved the abstraction of well-established sonification strategies "based on how many data points are rendered into the basic time interval, how many data dimensions are being used in the representation, and how many perceptual streams are in use". To that effect, audification, parameter mapping and model-based sonification methods were redefined as continuous, discrete-point, and model-based data representations, respectively.

It is important to note that these representations are guided by the fundamental notion of a time frame for the formation and attention of auditory gestalts:

> "In auditory gestalts (or sound objects) of 100 milliseconds and less it becomes more and more difficult to discern meaningful detail, while following a single gestalt for longer than say 30 seconds is nearly impossible, or at least takes enormous concentration; thus, a reasonable rule of thumb for single gestalts is to time-scale their rendering into the duration of echoic memory and short term memory, i.e. on the order of 1-3 seconds."

While this specifies a period in which data should be represented, the various representations indicate there is a great deal of flexibility in the number of data points involved. For instance, granular synthesis relies on a high density of short audio events which, via the process of sonification, can derive from an equally high number of data events. Conversely, a less complex tone could express a single data point mapped only to its pitch. In essence, both the definite time frame and the adjustable quantity of data points can be regarded as essential for underpinning the foundation for data transformation.

When discussing the visual interpretation of rigid body data we established that data values inform the state of a visible mode. This setup is not only indicative of a MBS methodology but can also prove effective for discerning and analysing auditory gestalts. To that effect this sonification design permits the allocation of one voice per data set which is manifest as a single rigid body, or object. Allocating a subsequent voice to the same object ensures that the previous instance ceases to exist in the same manner that is typified by a monophonic instrument. Each voice is controlled by an ADSR envelope which remains at the sustain stage while the conditions for the existence of this voice are met. The immediate reasoning behind this is to counteract extraneous clicks caused by an abrupt onset, or offset, of audio while adhering to a standard amongst hardware and software synthesis implementations. However, this practice also guarantees that an object will only generate a single sound stream for any given moment in time which can assist the listener in identifying each object to sound relationship, particularly if distinct timbres are used. De Campo observed that "when the dimensions in a data set are directly comparable it is conceptually convincing to render them as parallel streams". Every rigid body in a physics engine is described from another instance of the same data set, with more complex derivatives adding varying levels of abstraction. As such, each object can produce an audio stream that is directly comparable to another object's stream where multiple instances will generate parallel streams. When recalling the implications of auditory scene analysis we can accept that parallel streams of generated sound may fuse or separate based on a perceptual context. In the context of this framework the user can rely on the auxiliary feedback of the visual channel to assist comparisons as to what fuses, segregates or has no discernible effect.

Although it is apparent that MBS has distinct qualities which support the search for auditory gestalts, De Campo reasoned that "assumptions built into models may introduce bias leading away from understanding the domain at hand." One important assumption made by the MBS approach is that sound is generated in response to user interaction. Under typical circumstances rigid bodies in a simulation will act in consequence of the calculations performed by the computer. Dynamic bodies in particular will continue to act after long after any form of human interaction has set in place a sequence of events. During the course of the dynamic body's actions, its data values are continuously informed by the environment, along with other bodies, to produce a wealth of context sensitive information. When factoring in some key elements from a discrete point data representation, or parameter mapping sonification (PMS), this data can be directly mapped to sound in a more explicit manner that no longer overlooks the absence of user interaction with the model. De Campo distinguished this representation as follows:

> "Discrete Point Data Representation creates individual events for every data point. Here, one can easily arrange the data in different orders, choose subsets based on special criteria (e.g. based on navigation input), and when special conditions arise, they can be expressed well."

By applying this approach to each model the parameters for their voice can be continually controlled by a series of events which derive from their corresponding data set. This is particularly salient when considering that the data set itself is ideally suited for an event based method due to the manner in which it is generated. De Campo acknowledges that "mapping data time to listening time is metaphorically very close and thus easy to understand". This was echoed by Hunt and Hermann [22] who stated in their assessment of interactive PMS design that "if the data are themselves time-stamped, it is straightforward to map the time value onto the sonification time". A typical physics engine will generate a new set of parameters for each rigid body every time the engine is updated which, in turn, can be denoted by the designer. In order to achieve a sense of realism the value of this time interval is generally low enough to incur updates at a frequency that either matches or exceeds the visual frame rate. Nonetheless, it is unfeasible to suggest that this data rate could match that of the high audio sampling rate recommended for a continuous data representation as the demands it places on the processor would make real-time interaction impossible. Instead, a discrete point data representation provides a configurable arrangement that can stream parameters at a rate in keeping with the physics engine while effectuating an interactive and dynamic audiovisual environment. It also offers explicit control over the dimensions of each data set, making it possible to define a mapping configuration which involves any number of comparable attributes, or subsets, while maintaining that omitted parameters have no influence over the audio. In that respect, the user is given the opportunity to observe de Campo's recommendation that "generally, when multiple streams are used in a sonification design, the individual streams can and should use fewer dimensions."

Moody et al. [23] accepted that "our experience with objects which do emit sound when we can see they're in motion seems to make our brain much more receptive to linking otherwise unconnected audio and visual phenomena when it perceives a certain similarity in the temporal information of the two streams". One fundamental aspect of this foundation for data transformation is that both the audio and visual data representations will evolve simultaneously as they both rely on the values obtained from the periodic updates of the physics engine. This creates an ideal scenario that can be configured to benefit from our heightened perceptivity. For instance, the speed parameter, calculated from the normalisation of the linear velocity, will inform the model to visually describe the motion of the object, while mapping speed to the amplitude of the object's voice will allow the same parameter to simultaneously act as sonic metaphor for the same motion. While this simultaneous monitoring of sound and audio can be beneficial, Serafin et al. [24] warned that "there may be sound categories with very salient sonic parameters which are perhaps very intuitive, yet the sound would be less pleasant for long-term use, or even irritating or provoking an unwanted emotional reaction". A lengthy time frame also presents an issue for the attendance of auditory gestalt as it demands an unfeasibly high level of concentration from the user. This can be compensated for within the current mapping approach. By mapping speed to amplitude the user is able to control the length of each voice through a motion based metaphor which, at the same time, highlights the number of active parallel streams.

When incorporating a parameter mapping approach to sonification each data set can become sensitive to special conditions that delegate the flow of information and determine the presence of sound without being solely dependent on, or predicated by, human interaction. By assigning a condition to an interactive model the user is provided with a distinct mechanism over the timing of the rigid body's voice and, in consequence, a period in which to monitor and evaluate the data. This is guaranteed by the envelope accompanying each voice as it maintains a definite level of control that is independent of any mapping configuration. So far only one condition has been described which assumes the continuous presence of a voice to be characterised by any combination of data parameters via a mapping configuration. With the introduction of a proximity condition the user has a means of punctuating continuous audio on the grounds that each voice will only remain active while its corresponding rigid body is in close proximity to another rigid body. In this sense, although the voice can be regarded as continuous while the condition is met, the user now has a greater degree of control over which data sets will meet this condition. Moreover, the visually apparent nature of each model facilitates the viewer's perception of the spatial relationship between rigid bodies where the automatic sorting and analysis of their data sets can be supported by the gestalt law of grouping by proximity.

While the proximity condition admits a certain level of control over an otherwise continuous signal, the signal still lends itself to a prolonged time frame. However, in the interest of a more concise period of data exploration we can refer to our understanding that sound generally accompanies interaction between two or more rigid bodies. At its most fundamental level this interaction can be described as an impact, or collision. This was understood by Gaver [18] when he put forward that:

> "impacts are a basic-level event in the sense that they are produced by a simple interaction of objects; combinations of impacts may produce more complex events such as footsteps, hammering, or bouncing noises. Because they are basic-level events, understanding the information they convey is useful in understanding a great many more complicated events."

The same notion can apply to physics engines where interactions are described through a series of impulses [25]. Typically, the duration of each impulse is so brief that any corresponding voice would be hard to distinguish as its time frame would fall short of the criterion for the formation of an auditory gestalt. For this reason, the collision condition allows the voice's envelope to be configured by the user, where its overall length is bound by the same criterion. Each envelope operates as a dynamic tool for encapsulating the acoustic consequence of the kinetic events which, in turn, signify impulsive exchanges of momentum and energy within the simulation. These exchanges can be considered significant as they represent a discontinuity in the data which is discernible from the visual behaviour of the rigid bodies involved. This behaviour can also be supported by the gestalt principle of good continuation which specifies that abrupt changes along a mapped dimension are indicative of a notable event [20].

Levitin et al. [26] identified two distinct scenarios for the manifestation of musical events they affirmed that our auditory perception is consistent with this principle. The first of these scenarios, termed as an explicit beginning, stemmed from the listener's perception of an intensity discontinuity whereas the second, referred to as an implicit beginning, was formed on the basis of a perceived spectral discontinuity. Based on their findings it is evident that the user should be acutely aware of this principle should they wish to modify either the intensity or the spectral properties of a voice. This should apply to all conditions in which a voice is present and not just to the collision condition described here. In that respect, any mapping function that allows for an erratic fluctuation in the value of an audio dimension could unintentionally encourage the formation of an auditory gestalt and serve to mislead the listener, particularly if the underlying data suggests nothing of interest.

Voice Condition	Description
Continuous	Voice is continuously active
Proximity	Voice is active while its parent body is within a given proximity of another body
Collision	Voice is active at the onset of a collision involving its parent body (duration is denoted by an envelope)

Table 1: Conditions for a rigid body to be allocated a voice

The data transformation framework currently presents the user with a choice of three conditions to determine the existence of a voice (Table 1). While all conditions denote a loose time frame for the listener to attend to the audio, their current stage of implementation assumes that the data is to be constantly streamed within this period. In turn, this indicates that a mapped data parameter will continuously update its corresponding voice parameter while the voice is active. As previously established with amplitude, there are voice parameters which may require dynamic control throughout the duration of the envelope. Conversely, dynamic control may not be appropriate for other audio dimensions such as frequency, where the user might wish to generate a series of stable and discrete tones in keeping with traditional Western music. For this purpose, a single value sent once at the onset of each voice presents a viable alternative. Although this streaming condition treats a parameter as static, the underlying data remains dynamic. The static value therefore derives from the current state of the dynamic data each time the voice condition is met. For example, every time an object is involved in a collision it will generate a new voice instance that is informed by the current static state of the dynamic data. In this case, a new collision event would indicate a potential change in frequency. From a creative standpoint each collision can now generate enough variety to be considered musically dynamic despite the frequency being more discrete and controlled. From a more analytical perspective the static value can also be sustained by the voice parameter. This makes it ideal to obtain an auditory snapshot of the data parameter at the time of the event. However, it is worth considering that the condition most appropriate for either scientific analysis or musical ambition will be dependent on both the parameter and the preference of the user. For that reason each data streaming condition (Table 2) can be independently applied to a parameter in order to determine its rate of transfer.

Data Streaming Condition	Description
Voice Synchronous	Send data while the voice is active (data is sent each time the physics engine is updated)
Voice Onset	Send data only once at the moment the voice becomes active

Table 2: Conditions for the streaming of data from a rigid body

3.3. Summary

This section has described a foundation for data transformation (Figure 2) which is supported by a hybrid sonification method known as model-induced parameter mapping [22]. To better understand why this particular approach has been chosen it will now be compared to a more established process for generating the sonic counterpart of a physically simulated object. By demonstrating a fundamental difference with respect to the connection between data and sound it can be argued that the sonification method presented here holds some key advantages.

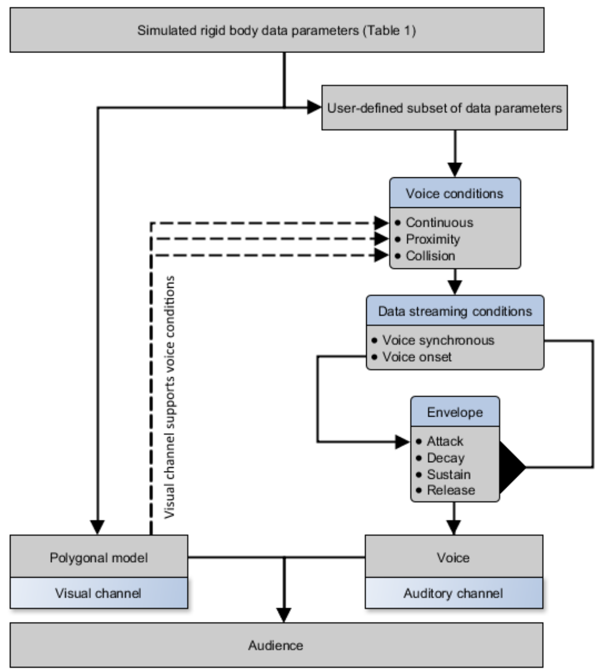

Figure 2: Diagram explaining the model-induced parameter mapping approach used for sonifying the flow of rigid body data

4. ADVANTAGES OVER THE PHYSICAL MODELLING APPROACH

A physical model uses mathematical formalisms to simulate the sound source of an object based on the understanding and implementation of its sound production mechanism. Accordingly, a computer can be used to generate convincing replications of existing sources, such as musical instruments, while presenting the means to extend their sonic functionality beyond the limitations of the physical world [13]. Although this methodology can be considered to provide an intuitive link between data and sound, it is perhaps restricted by its imitative nature.

For both physical modelling and PMS, synthesis is derived from the internal state, or set of parameters, that describe the object. However, the physical modelling approach differs in the fact that both the mechanical and acoustical systems are governed by classical physical laws which denote the intrinsic link between motion and sound. Cadoz [27] explained that, with a mass-interaction approach to physical modelling, mechanical movements, or vibration modes, must be constrained to exclusive frequency bandwidths in order to separate audible objects from their non-audible equivalents. By contrast, any rigid body using this framework can be a source of energy for performing an action, or triggering a chain of events, while providing the user with full control over their sonic activity. This implies a means for a greater potential of gestural interaction while circumventing the need for some imposing restrictions. For instance, Cadoz uses the metaphor of the maraca for the simulation of free particles in a box. This could be seen as a restriction brought about by the tight coupling of the two systems where the number of vibrating bodies involved alludes to the generation of noise.

Gaver [18] identifies that there is a one-to-many mapping between the sound and the parameter under consideration, which he perceives as a flaw:

> "A change in an object's length produces a change of the fundamental frequency of the sound it makes. But other attributes of an object, such as its shape, density, and hardness, also determine its functional frequency. Thus, an impact sound's fundamental frequency does not specify its length."

The resultant sound is therefore unable to reveal particular information about the event without the use of specified constraints, that is, all other attributes must remain constant. With that in mind, Ma et al. [28] found that "similar materials or textures of the sounds sources, similar effects of the interactions, and similar events that take place can all be the cause of confusion". In order to better facilitate the communication of ideas they proffer that "distinctive environmental sounds can effectively evoke concepts (nouns and verbs) commonly used in everyday communication". Indeed, physically modelled objects that have a large number of parameters in common can produce indistinguishable sounds. This is particularly true when the sound is modelled from a limited subset of data, such as those pertaining to the contact point [29]. However, these restrictions need not apply to the synthesis routines associated with PMS as they are not determined by the resonating structures themselves but by sonifying any combination of the parameters that attribute to the simulated object's behaviour. In other words, the user can determine the number of rigid body properties that are mapped to sound and thus the level of sonic complexity over which they have control. Ergo, when recalling Gaver's object to sound correspondence it is possible to argue that an impact sound's fundamental frequency can specify the object length should the user choose to link that particular attribute to the relevant sound dimension. Furthermore, given the plethora of available synthesisers, along with their wide range of sounds, the user is presented with a rich palette for communication. This includes the fulfilment of more environmentally accurate material sounds when sending the underlying data parameters to those of a physical modelling synthesiser.

In the same way that electronic musical instruments physically decouple the control interface from the sound generator a PMS would furnish a decoupling of the sound from the acoustic properties of a simulated object. By adopting this alternative approach we are presented with a visual metaphor that fits our everyday observations but the sound representation has become more subjective. As such, the user must have greater involvement in understanding how the data dimensions should be linked to the audio dimensions. Menzies [29] regarded physical modelling as a limiter of expression stating that "explicit physical models are often difficult to calibrate to a desired sound behaviour although they are controlled directly by physical parameters". He advocates that "the sound designer is often more interested in the freedom to shape the sound how they would like, rather than exactly matching a real behaviour that may not be quite suitable". In contrast, the PMS approach used in this framework provides a more explicit link between data and sound where the user is granted with full control over each stage of the transformation process in order to facilitate a greater capacity for expression. This follows the advice given by Hunt and Wanderley [30] who suggested that "explicitly defined mapping strategies present the advantage of keeping the designer in control of the design of each of the instrument's component parts, therefore providing an understanding of the effectiveness of mapping choices in each context".

5. CONCLUSION

This investigation has outlined an alternative approach for the interpretation of simulated rigid body data. By employing a hybrid sonification method known as model-induced parameter mapping, the user can rely on both the audio and visual channels to interpret the same data. While the visual channel presents the data in the form of an intuitive model form that relates to our common understanding for the comprehension of tasks and events, it also provides auxiliary feedback for the coexisting auditory channel. However, although this auditory channel is informed by the same data set as the visual channel, it abandons the strict simulation of physical ties between mechanical and acoustic systems in favour of a PMS methodology that is informed by several significant modes of listening. This provides the scope for the characterisation of particular voice conditions where the arrangement and classification of information flow is supported by gestalt laws, the emergence of which can be predicated and reinforced by the model. Furthermore, as each model enforces monophonic behaviour they are guaranteed to be associated with no more than a singular voice, or sound stream, at any given moment. The direct comparison of rigid body data is now feasible since each additional model is capable of representing a single audible stream which derives from a unique instance of the same pool of parameter types. Consequently, the state of the model can also support the perceptual organisation of these parallel streams by depicting the emergence of grouping cues familiar to auditory scene analysis.

When evaluating this data transformation framework in the context of an auditory display it was shown that the user gains a number of advantages over utilising a physical modelling approach. In particular, the method presented here creates a more explicit link between data and sound which circumvents the need to empirically calibrate parameters in an indirect manner in order to change a particular quality of sound. This allows for the sonic analysis of any user-defined combination of rigid body data parameters and a greater choice of sound synthesis routines to describe them.

6. FUTURE WORK

Although this framework supports an explicit link between rigid body data and sound, the freedom of control associated with the inclusion of PMS raises concerns over an issue known as "The Mapping Problem" [31]. Initial investigations have been made which address this issue from the perspective of rigid body dynamics [6]. However, this is subject to ongoing research which cannot be covered in any depth within the scope of this paper.

7. REFERENCES

[1] http://www.unity3d.com

[2] D. B. Lloyd, N. Raghuvanshi and N. K. Govindaraju, "Sound synthesis for impact sounds in video games," in *Symposium on Interactive 3D Graphics and Games*, 2011, pp. 55-62.

[3] J. F. O'Brien, C. Shen, and C. M. Gatchalian, "Synthesizing sounds from rigid-body simulations," in *ACM SIGGRAPH Symposium on Computer Animation*, 2002, pp. 175-181.

[4] A. Dolphin, "Compositional applications of a game engine. Creative practical applications in sound art and music composition," in *Proceedings of the Games Innovations Conference*, 2009, pp. 213-222.

[5] R. Hamilton, "Q3OSC or: how I learned to stop worrying and love the ~~bomb~~ game," in *Proceedings of the International Computer Music Conference*, 2008.

[6] R. Perkins, "Interactive sonification of a physics engine," PhD thesis, Anglia Ruskin University, Cambridge, 2014.

[7] http://www.rhysperkins.com/?works=mhyusics

[8] D. Worrall, "Sonification and information concepts and techniques," PhD thesis, University of Canberra, Australia, 2009.

[9] A. Hunt and T. Hermann "The importance of interaction in sonification," in *Proceedings of the International Conference on Auditory Display*, 2004, pp. 1-8.

[10] S. Barrass, (1998) "Auditory information design," PhD thesis, The Australian National University, Australia, 1998.

[11] A. F. Blackwell, "The reification of metaphor as a design tool," *ACM Transactions on Computer-Human Interaction*, vol. 13, no. 4, 2006, pp. 490-530.

[12] A. Pirhonen, "To simulate or stimulate? In search of the power of metaphor in design," in *Future Interaction Design*, A. Pirhonen, H. Isomäki, C. Roast and P. Saariluoma eds., 2005, London, UK: Springer Verlag, pp. 105-123.

[13] C. Cadoz, "Supra-instrumental interactions and gestures," *Journal of New Music Research*, vol. 38, no. 3, 2009, pp. 215-230.

[14] J. Kojs, "The language of action and cyberaction-based music: theory and practice," *Journal of New Music Research*, vol. 38, no. 3, 2009, pp. 285-294.

[15] N. Castagne, C. Cadoz, A. Allaoui, and O. Tache, "G3: GENESIS software environment update," in *Proceedings of the International Computer Music Conference*, 2009, pp. 407-410.

[16] T. Hermann, (2002) "Sonification for exploratory data analysis," PhD thesis, Bielefeld University, Germany, 2002

[17] T. Hermann, and H. Ritter, "Sound and meaning in auditory data display," *Proceedings of the IEEE*, vol. 92 no. 4, 2004, pp. 730-741.

[18] W. W. Gaver, "How do we hear in the world? Explorations in ecological acoustics," *Ecological Psychology*, vol. 5, no. 3, 1993, pp. 285-313.

[19] G. Kramer, "An introduction to auditory display," in *Auditory Display: Sonification, Audification, and Auditory Interfaces*, G. Kramer, ed., 1994, Reading MA: Addison-Wesley, pp. 1-77.

[20] A. S. Bregman, *Auditory scene analysis: the perceptual organization of sound*, 1990, Cambridge, MA: Bradford Books, MIT Press.

[21] A. de Campo, "Science by ear. An interdisciplinary approach to sonifying scientific data," PhD thesis, University for Music and Dramatic Arts Graz, Austria, 2009.

[22] T. Hermann, (2011) "Model-based sonification," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, eds., 2011, Berlin, Germany: Logos Publishing House, pp. 399-427.

[23] N. Moody, N. Fells, and N. Bailey, "Motion as the connection between audio and visuals," in *Proceedings of the International Computer Music Conference*, 2006, pp. 405-408.

[24] S. Serafin, K. Franinović, T. Hermann, L. Guillaume, M. Rinott, and D. Rocchesso, "Sonic interaction design," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, eds., 2011, Berlin, Germany: Logos Publishing House, pp. 87-110.

[25] I. Millington, *Game Physics: Engine Development*, Series in Interactive 3D Technology, San Francisco, CA: Morgan Kaufmann, 2007, p. xx.

[26] D. J. Levitin, S. McAdams and R. L. Adams, "Control parameters for musical instruments: a foundation for new mappings of gesture to sound," *Organised Sound*, vol. 7 no. 2, 2002, pp. 171-189.

[27] C. Cadoz, "The physical model as metaphor for musical creation "pico..TERA", a piece entirely generated by physical model," in *Proceedings of the International Computer Music Conference*, 2002, pp. 305-312.

[28] X. Ma, C. Fellbaum and P. R. Cook, "Environmental sounds as concept carriers for communication," in *Proceedings of the 16th International Conference on Auditory Display*, 2010, pp. 353-360.

[29] D. Menzies, "Phya and virtual foley, physically motivated environmental sound synthesis for virtual worlds," *EURASIP Journal on Audio, Speech and Music Processing*, 2010, pp. 1-11.

[30] A. Hunt and M. M. Wanderley, "Mapping performer parameters to synthesis engines," *Organised Sound*, vol. 7 no. 2, 2002, pp. 97-108.

[31] D. Worrall, "A method for developing an improved mapping model for data sonification," in *Proceedings of the International Conference on Auditory Display*, 2011.